

Sentimen Etika Posting Pada Media Sosial Menggunakan Performa Terbaik

Ridwansyah^{1*}, Sopian Aji²,

^{1*} Teknik Informatika; STMIK Nusa Mandiri Jakarta; Jl. Damai No. 8 Warung Jati Barat Margasatwa Jakarta Selatan. Telp (021)-78839513/Fax (021) 78839421; e-mail: ridwansyah.rid@bsi.ac.id

² Sistem Informasi; STMIK Nusa Mandiri Jakarta; Jl. Damai No. 8 Warung Jati Barat Margasatwa Jakarta Selatan. Telp (021)-78839513/Fax (021) 78839421; e-mail: sopian.sop@bsi.ac.id

* Korespondensi: e-mail: ridwansyah.rid@bsi.ac.id

Diterima: 21 November 2017; Review: 22 November 2017; Disetujui: 24 November 2017

Cara Sitasi: Ridwansyah, Sopian Aji. 2017. Sistem Etika Posting Pada Media Sosial Menggunakan Performa Terbaik. Information Management For Educators And Professionals. 2(1): 67-76.

Abstrak: Analisis sentimen berisikan dataset berbentuk teks yang bersifat positif, negatif atau netral dengan mengungkapkan kata-kata dalam media sosial haruslah berhati-hati dan jangan ada unsur menghina atau melakukan pelanggaran hukum dapat mengakibatkan terkena pidana bagi mereka yang melakukan pelanggaran hukum terkait postingan di media sosial. Metode klasifikasi seperti *Naive Bayes*(NB) diusulkan oleh banyak peneliti untuk digunakan pada analisis sentimen teks. Dari metode naive bayes, akan diuji dengan tiga masukan yaitu menggunakan *Tokenize* (menghilangkan tanda baca), *Transform Cases* (merubah huruf besar menjadi huruf kecil) dan *Stopwords* (menghilangkan kata negatif) dari komentar bersifat positif (100 komentar teks) dan bersifat negatif (100 komentar teks). Dari hasil penelitian dengan menggunakan *Naive Bayes* didapat nilai *AUC*: 0.484 +/- 0.107 (mikro: 0.484) (positive class: *Positif*) dengan hasil akurasi 72,50, maka metode *Naive Bayes* dapat menghasilkan klasifikasi yang cukup untuk analisis sentimen etika posting pada media sosial.

Kata kunci: Etika Posting, *Naive Bayes*, Media Sosial.

Abstract: Sentiment analysis with a unique dataset shaped text is positive, negative or neutral with expressing words in media social shall be careful and follow the insult or done can result in exposed for those who commit criminal law violations related to post on social media. A method of classification as naive bayes(nb) proposed by many researchers for use in the analysis sentiment text. Of the bayes naive, will be tested with three input that is using tokenize (remove) punctuation, the cases transform instead (change and a small type) stopwords (took the word negative) (100 of commentary is positive and negative comments text) (100) text commentary. The research using bayes naive obtained the auc: 0.484 + / - 0.107 micro (:0.484) (:class it with the positive) 72,50 accuracy, hence a method bayes can produce classifications naive enough to analysis.

Keywords: Ethics post, *Naive Bayes*, Social Media.

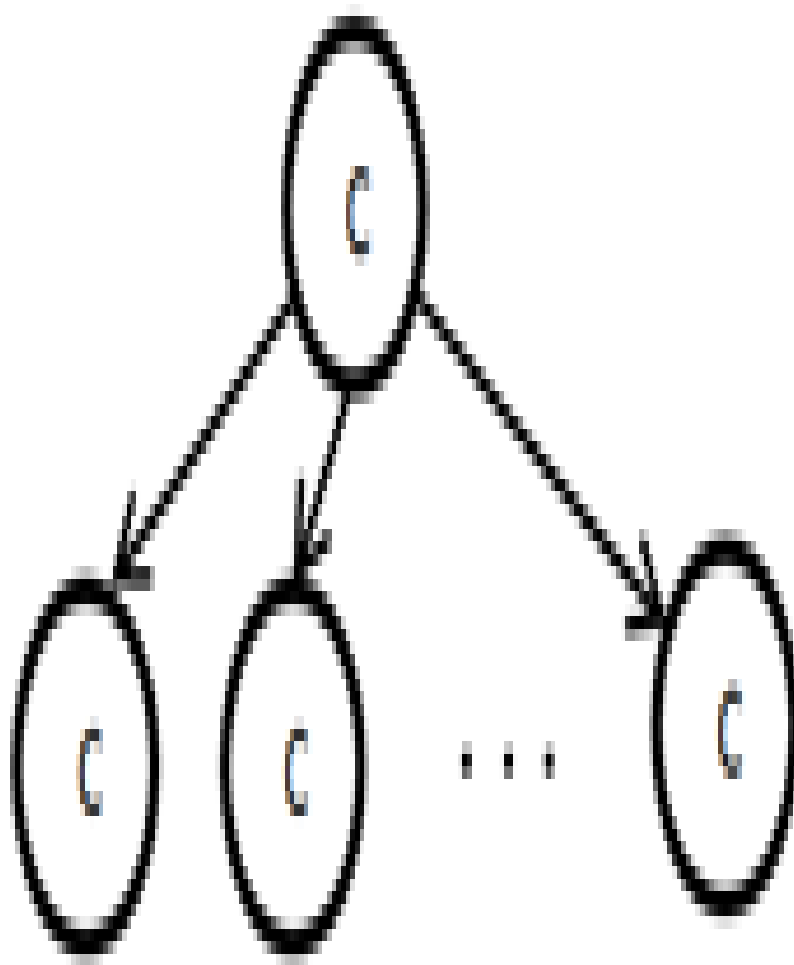
1. Pendahuluan

Informasi yang diperoleh dari media Televisi, Surat Kabar, Majalah, Radio atau Internet sangat dibutuhkan oleh masyarakat dengan berbagai macam aspek yang dibutuhkan dapat berupa dokumen teks yang dapat diakses melalui halaman web, internet terus menjadi bagian penting dari kehidupan sehari-hari. Masyarakat sekarang mampu mengakses internet tidak hanya dari anggota keluarga dan teman, tetapi juga dari orang asing yang terletak di seluruh dunia yang dapat menggunakan sebuah produk tertentu, mengunjungi sebuah tujuan tertentu, atau berita, namun sekarang ini berbagai media sosial yang ada mulai menggunakan sistem berbasis web, Media Jejaring Sosial terdiri dari berbagai media online seperti blog, forum, review produk, game sosial, *Path, Line, Instagram, Whatsapp, Blackberry Messenger, Twitter, Facebook, Wikipedia, Google Plus dan Youtube*.

Pada umumnya informasi berita yang disampaikan bersifat positif tetapi ada juga yang bersifat negatif, orang sering mengungkapkan sentimen atas perasaan yang dirasakan mereka melalui media sosial yang memfokuskan perhatian pada pendapat sentimen, perasaan atau emosi. Kata kata 'sentimen' dan 'pendapat' yang sering digunakan secara bergantian. Sentimen bisa digambarkan sebagai emosi, keputusan, pendapat atau ide-ide. Postingan internet harus mempunyai beberapa etika dalam mengomentari sebuah forum atau review dan lain-lain, karena berbagai macam informasi yang didapatkan secara mudah oleh masyarakat. Pemberitaan yang negatif pada media sosial akan sangat berbahaya karena semua kalangan dapat mengakses informasi tersebut misalnya dari segi usia, golongan, suku, agama bahkan antar-negara, contoh dari pemberitaan yang bersifat negatif di media jejaring sosial telah banyak beredar kasus pornografi, penghinaan, pencemaran nama baik, penistaan, perbuatan tidak menyenangkan, memprovokasi, menghasut, menyebarkan berita bohong dan semua tindakan memiliki tujuan atau bisa berdampak pada tindak diskriminasi, kekerasan, penghilangan nyawa dan konflik sosial.

Sentimen terutama mengacu pada perasaan, emosi, pendapat atau sikap. Dengan meningkatnya *World Wide Web*, orang sering mengungkapkan perasaan mereka melalui internet di media sosial, blog, rating dan reviews. Karena ini meningkatkan data tekstual, ada kebutuhan untuk menganalisa konsep mengekspresikan sentimen dan menghitung wawasan untuk menjelajahi bisnis [Tripathy et al., 2015]. Analisis sentimen, adalah proses bertujuan untuk menentukan apakah polaritas korpus tekstual (dokumen, kalimat, paragraf dll) cenderung ke arah positif, negatif atau netral. [Kontopoulos et al., 2013]. Opini mining tidak memperhatikan topik dari teks tersebut tetapi lebih fokus kepada ekspresi yang digambarkan dari teks opini tersebut. Hal ini menentukan komentar dalam forum online, blog, atau komentar yang berkaitan dengan topik tertentu (produk, buku, film dan lain-lain) termasuk opini positif, negatif atau netral [Martinez et al., 2014]. Berbagai pendekatan pembelajaran mesin untuk otomatisasi klasifikasi sentimen, misalnya Teori *Naïve Bayes*, *Maximum Entropy (ME)*, *Support Vector Machine (SVM)*, *Decision Tree* maupun *Neural Network*.

Naive Bayes Classifier (NBC) adalah metode sederhana tetapi memiliki akurasi serta performansi yang tinggi dalam pengklasifikasian teks [Routray et al., 2013]. Prediksi dari suatu anggota probabilitas dengan klasifikasi statistik merupakan klasifikasi-klasifikasi *naive bayes*. Dalam *Naive Bayes* diasumsikan prediksi atribut adalah tidak tergantung pada kelas atau tidak dipengaruhi atribut lain. [Darujati and Gumelar, 2012]. Bentuk *naive bayes* yang paling sederhana dari bayes adalah jaringan, dimana semua atribut diberikan nilai independen dari kelas variabel. Untuk mengatasi keterbatasan ukuran *naive bayes* adalah strukturnya untuk mewakili secara eksplisit yang ketergantungan diantara atribut. Suatu desain klasifikasi dengan memisah objek dengan masalah yang ada dengan sifat hipotesis merupakan ide dasar dari *naive bayes* dengan pendekatan statistik dalam pengenalan pola dengan nilai atribut secara kondisional saling tidak tergantung jika diberikan nilai *output*. Metode *Naive bayes* dalam proses klasifikasi teks mempunyai tahap pelatihan dan tahap klasifikasi dengan dilakukan proses analisis sampel dokumen dengan kata yang sering tampil dalam koleksi dokumen yang menjadi representasi dokumen. Setelah itu penentuan probabilitas tiap kategori berdasarkan sampel. Dengan gambar sebagai berikut:



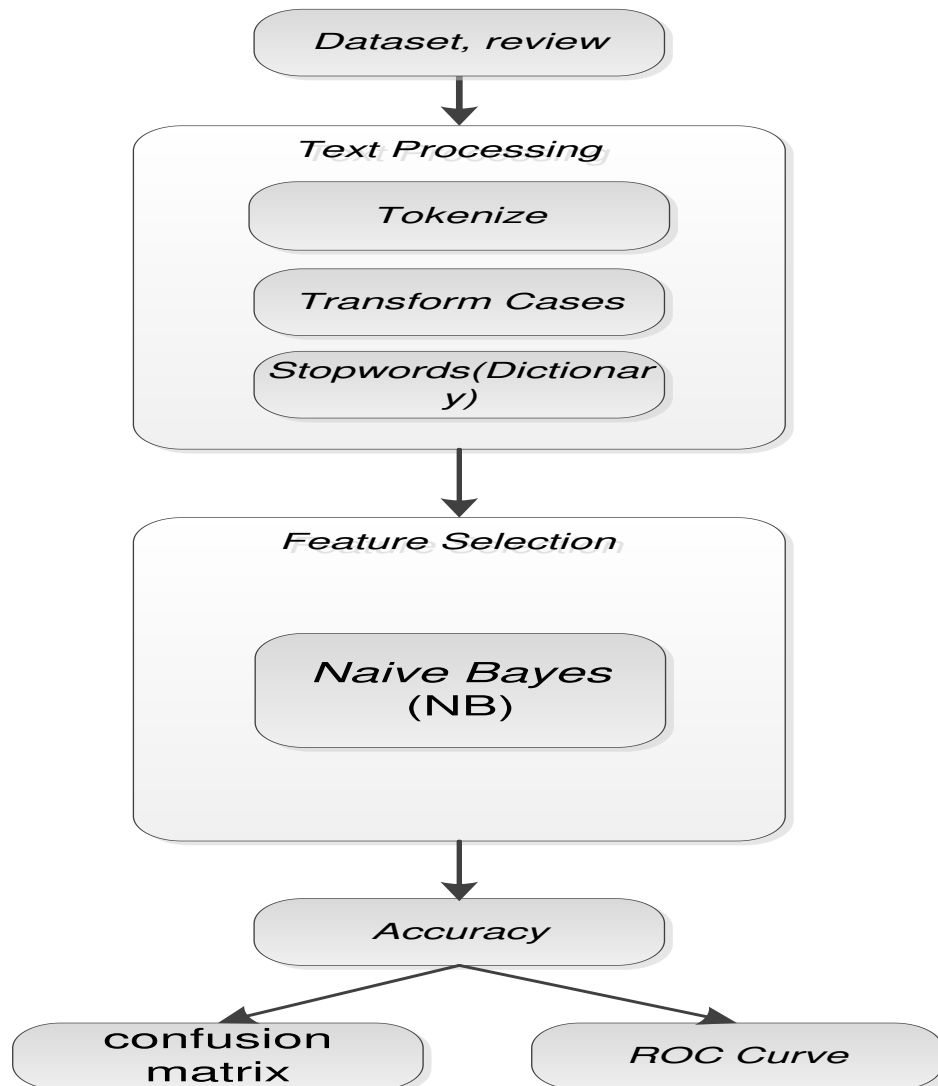
Sumber: Darujati and Gumelar (2012)

Gambar 1. Klasifikasi Naïve Bayes dengan atribut prediksi

Klasifikasi sentimen secara otomatis mengelompokkan komentar dari pengguna menjadi opini positif ataupun opini negatif. [Zhang et al., 2011] Oleh sebab itu jumlah data tekstual yang membicarakan hal negatif sangat meningkat, maka perlu dilakukan analisis konsep eksperimen sentimen untuk memprediksi apakah sebuah data yang mewakili sentiment positif atau *negative*, jika negatif maka dihilangkan atau diganti. Dari permasalahan diatas maka Identifikasi permasalahan pada sentimen etika posting media sosial dimulai dari adanya permasalahan pada sentimen posting pada sebuah berita dengan banyaknya pemberitaan di media sosial yang bersifat negatif sehingga perlunya pembatasan atau penyaringan pada pemberitaan yang sesuai keinginan kita dalam waktu yang singkat dan bagaimana tingkat akurasi analisa sentimen yang dihasilkan algoritma *Naïve Bayes* dengan *text mining* permasalahan yang berkaitan dengan penganalisaan teks dengan jumlah yang besar, penemuan penggalian informasi dari suatu teks yang nantinya akan sangat berguna. Maka tujuan dari permasalahan yang ada membantu para pembaca atau penerima informasi yang lebih sehat dan bersifat positif.

2. Metode Penelitian

Metode penelitian dengan mengukur akurasi data eksperimen yang dilakukan dalam penelitian dengan menggambarkan kerangka pemikiran yang diusulkan dalam penelitian ini sebagai berikut.



Sumber: Hasil Penelitian (2017)

Gambar 2. Kerangka Pemikiran

Data-data yang didapat dikumpulkan dan semua kata di tiap data dihilangkan tanda baca atau simbol yang bukan huruf, dengan kata-kata yang tidak relevan akan diubah dan kata-kata yang kasar yang ada didalam dataset akan dihilangkan dengan menggunakan seleksi fitur menggunakan *naïve bayes* dengan menghasilkan hasil akurasi berupa *confusion matrix* dan kurva ROC. Analisis sentimen kinerja sistem biasanya sering dilakukan melalui eksperimen dan evaluasi sistem akurasi, untuk melakukan eksperimen peneliti menggunakan data komentar *review*. Maka didapat 200 data *review* yang terdiri dari 100 *review* positif dan 100 *review* negatif. Data *review* positif disatukan dalam folder dengan nama positif. Sedangkan data *review* negative disatukan penyimpanannya dalam folder dengan nama negatif. Tiap dokumen berekstensi .txt yang dapat dibuka menggunakan aplikasi *Notepad*.

Kemudian setelah mendapatkan data positif dan negatif dalam bentuk .txt dengan jumlah data yang sama, yaitu 100 *review* data positif maka yang bersifat negatif harus didapatkan 100 *review* data negatif, dan data akan diproses. Untuk *text processing* yang peneliti gunakan *Tokenize*, *Transform Cases*, *Stopwords(Dictionary)*. Pengujian yang dilakukan dengan memilih *feature selection Naive Bayes(NB)* dengan bertujuan untuk mendapatkan nilai *accuracy*. Dari nilai *accuracy* persamaan model *confusion matrix* proporsi jumlah nilai prediksi didapatkan dan ROC Curve sehingga kita bisa melihat grafik ROC dengan nilai AUC (Area Under Curve).

Untuk tabel kontingensi dalam perhitungan adalah pada tahap evaluasi sistem, yang diberikan terlihat pada tabel 1.

Tabel 1. Tabel Kontingensi Evaluasi Kinerja Sistem

#	Relevant	Not Relevant
Retrieved	True Positive (TP)	False Positive (FP)
Not Retrieved	False Negatif (FN)	True Negatif (TN)

Sumber: Ling et al (2014)

Akurasi adalah ketepatan dalam menilai suatu sistem untuk melakukan klasifikasi yang benar. Perhitungan untuk Presisi (P) Recall (R), dan akurasi dapat dikalkulasi sebagai berikut:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$\text{Akurasi} = \frac{TP + TN}{TP + FP + FN + TN}$$

Dengan persamaan model confusion matrix sebagai berikut:

Rumus mencari nilai akurasi (acc) dengan proporsi jumlah prediksi yang benar.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Rumus mencari sensitivity digunakan untuk membandingkan proporsi TP terhadap tupel yang positif.

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Rumus mencari specificity digunakan untuk membandingkan proporsi TN terhadap tupel yang negatif.

$$\text{Specificity} = \frac{TN}{TN + FP}$$

Rumus mencari PPV (Positive Predictive Value) adalah proporsi kasus dengan hasil diagnosa positif.

$$\text{ppv} = \frac{TP}{TP + FP}$$

Rumus mencari NPV (negative predictive value) adalah proporsi kasus dengan hasil diagnosa negatif.

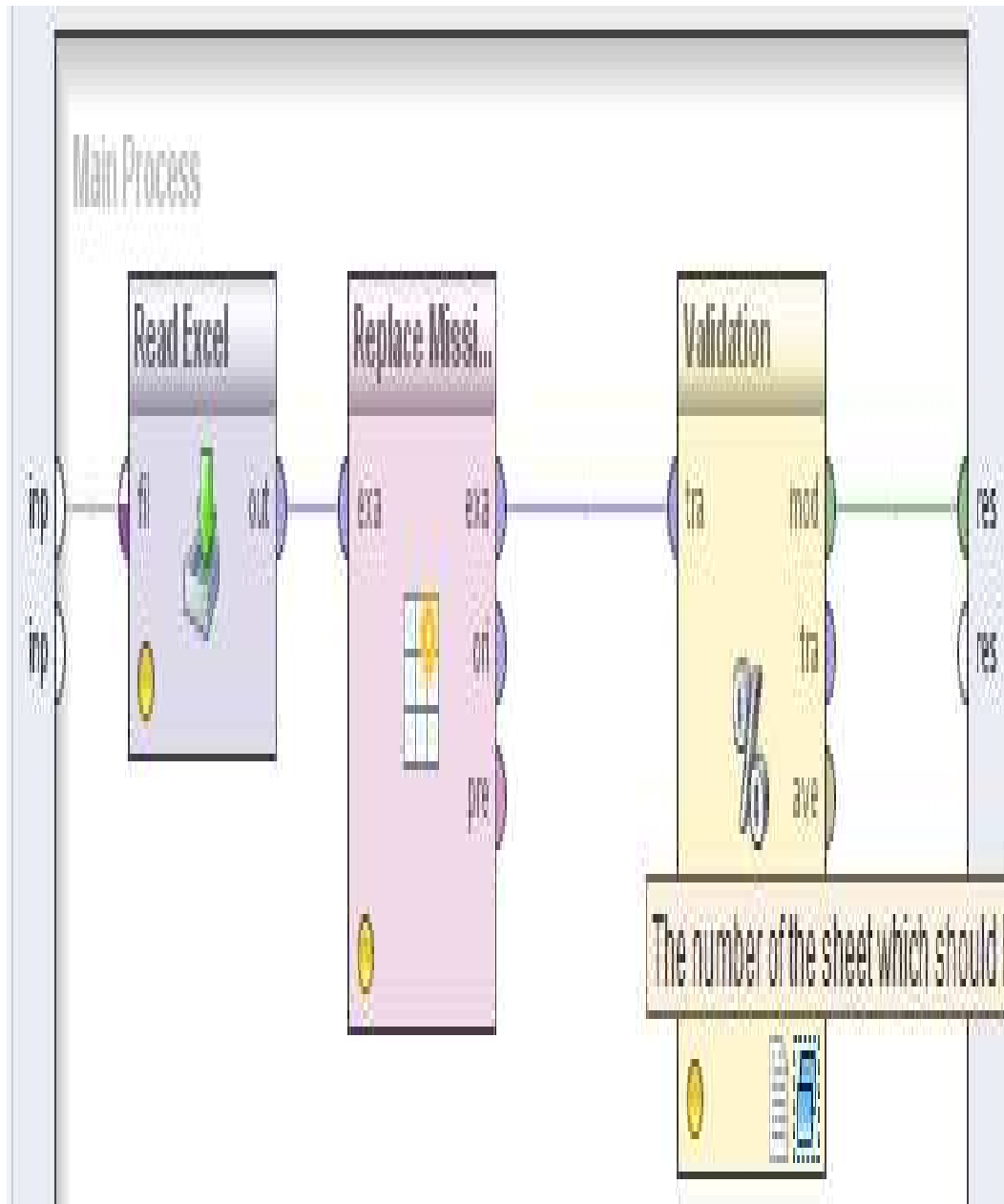
$$\text{npv} = \frac{TN}{TN + FN}$$

Metode penelitian yang penulis lakukan adalah metode penelitian eksperimen, dengan tahapan sebagai berikut: 1) Pengumpulan Data merupakan data untuk eksperimen ini dikumpulkan, lalu diseleksi dan dipisahkan dari data yang sesuai maupun yang tidak sesuai. 2) Pengolahan Awal Data merupakan model dipilih berdasarkan kesesuaian data dan permasalahan yang ada dengan metode yang baik dari beberapa metode pengklasifikasian teks yang sudah digunakan oleh peneliti sebelumnya dengan menggunakan metode naïve bayes. 3) Metode Yang Diusulkan merupakan metode yang diusulkan Untuk menghasilkan akurasi yang baik dari Algoritma *Naïve Bayes*. 4) Eksperimen dan Pengujian Metode merupakan untuk eksperimen data penelitian, penulis menggunakan RapidMiner untuk mengolah data. 5) Evaluasi dan Validasi Hasil merupakan evaluasi dilakukan untuk mengetahui akurasi dari model algoritma *Naïve Bayes* Validasi digunakan untuk melihat perbandingan hasil akurasi dari model yang digunakan dengan hasil yang telah ada sebelumnya. Teknik validasi yang digunakan adalah *cross validation*.

3. Hasil dan Pembahasan

Penelitian yang dilakukan menggunakan Aplikasi *RapidMiner Studio* dengan menggunakan spesifikasi komputer CPU *Intel Core i3 1.90GHz*, RAM 4GB, dan *sistem operasi Microsoft Windows 7 Professional 64-bit*. Hasil pengujian *review* dengan menggunakan teks bahasa Indonesia.

Pada penelitian sentimen etika posting menggunakan algoritma *naïve bayes* dengan menentukan nilai akurasi untuk menghasilkan suatu data yang berupa sentimen posting dengan desain model metode pengujiannya sebagai berikut:



Sumber: Hasil Penelitian(2017)

Gambar 3. Desain Model Validasi

Penelitian yang dilakukan dengan hasil dari penelitian menggunakan *Text Processing* pada table 2 yang dilalui terdiri dari *Tokenize*, *Transform Cases*, *Stopwords(Dictionar)*. Hasil pengolahan sebagai berikut:

Tabel 2. Text Processing menggunakan Tokenize, Transform Cases, Stopwords(Dictionary)

Text Review	Tokenize	Transform Cases	Stopwords(Dictionary)
Satu kata "LEBAY" kapan mau maju indo?? Klo peraturannya makin lama makin "GA PENTING"	Satu kata LEBAY kapan mau maju indo Klo peraturan nya makin lama makin GA PENTING	Satu kata lebay kapan mau maju indo Klo peraturannya makin lama makin ga penting	Satu kata kapan mau maju indo Klo peraturannya makin lama makin ga penting
SAATNYA JADI ORANG MUNAFIK,,MA RI BICARA YANG BAIK- BAIK MAU SUMPAAH SERAPAH CUKUP DALAM HATI AJA..wk..wk..	SAATNY A JADI ORANG MUNAFIK MARI BICARA YANG BAIK BAIK MAU SUMPAAH SERAPA H CUKUP DALAM HATI AJA wk wk	saatnya jadi orang munafikm mari bicara yang baik baik mau sumpah serapah cukup dalam hati aja wk wk	saatnya jadi orang mari bicara yang baik baik mau sumpah serapah cukup dalam hati aja wk wk
YESSS !!! PENJARA SEMAKIN RAMAI !!!	YESSS PENJAR A SEMAKIN RAMAI	yesss penjara semakin ramai makin	yesss semakin ramai makin
Kaya orang suci aja lu...BABI	Kaya orang suci aja lu BABI	kaya orang suci aja lu babi	kaya orang suci aja lu
MASIH MENDING W, DARI PADA LU BEGOOOO!!	MASIH MENDIN G W DARI PADA LU BEGOOO O	masih mending w dari pada lu begoooo	Masih mending w dari pada lu
Beritanya tidak bermutu FUCK!!!	Beritanya tidak bermutu FUCK	beritanya tidak bermutu fuck	beritanya tidak bermutu
Ahhh..cewek kayak lu kaya JABLAY :P	Ahhh cewek kayak lu kaya JABLAY	Ahhh cewek kayak lu kaya JABLAY	ahhh cewek kayak lu kaya

Sumber: Hasil Penelitian (2017)

Pada tabel 2 terlihat *Text Processing* dari *text review* yang bersifat negatif akan berubah menjadi *text review* yang bersifat positif. Pada *text review* melalui *Tokenize* seluruh tanda baca

akan dihilangkan, kemudian *Transform Cases* seluruh *text review* yang menggunakan huruf besar akan berubah menjadi huruf kecil dan yang terakhir *Stopwords(Dictionar)* kosa kata atau kata yang bersifat negatif akan dihilangkan.

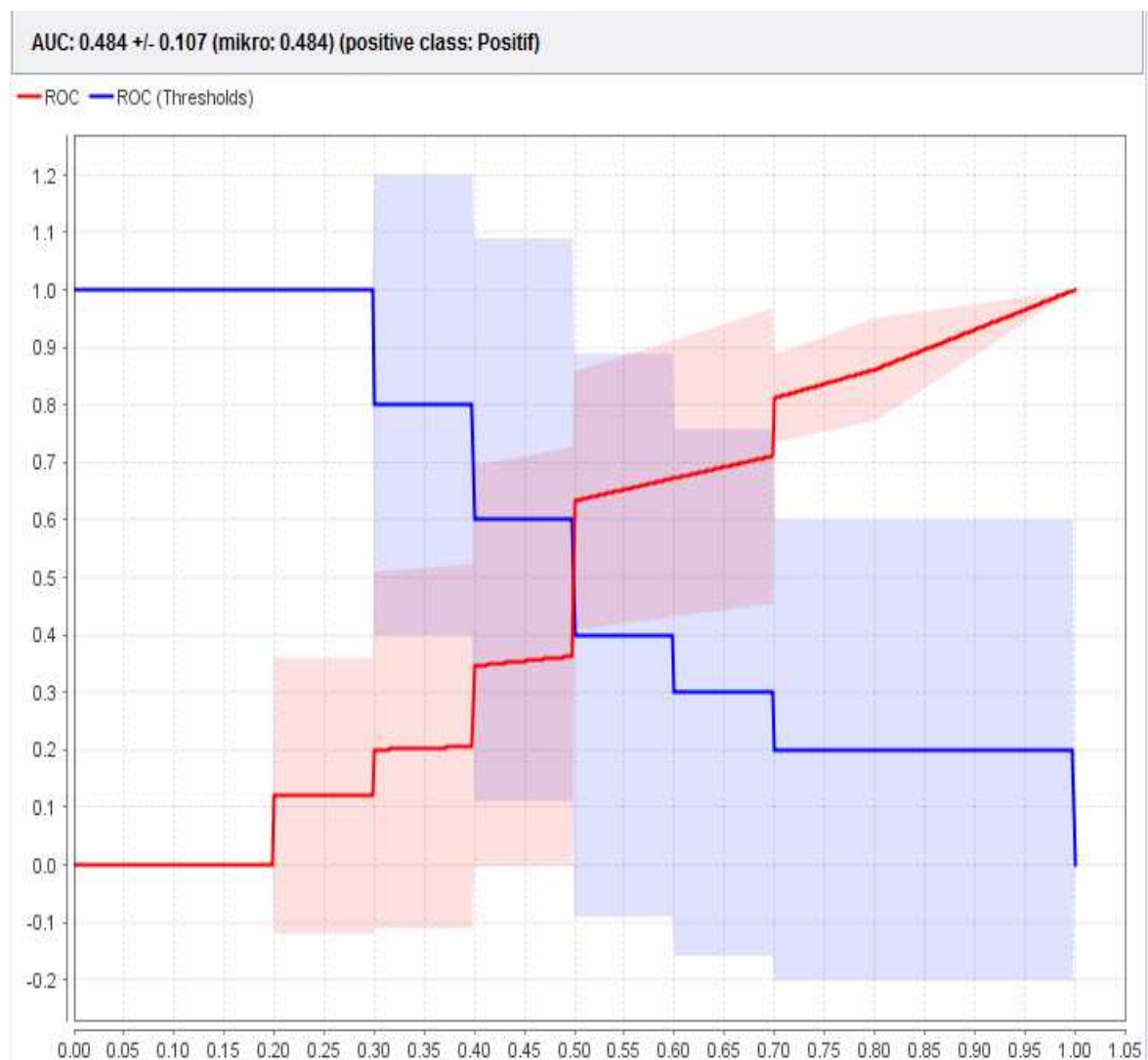
Pengujian penelitian dilakukan menggunakan algoritma *Naive Bayes* untuk mendapatkan *accuracy* dapat kita lihat pada tabel 3 dengan nilai *accuracy* 72.50%, sebagai berikut:

Tabel 3. Accuracy Algoritma Naive Bayes

	true Negatif	true Positif
pred. Negatif	58	33
pred. Positif	42	67

Sumber: Hasil Penelitian (2017)

Jumlah *true positive* (tp) adalah 58 *record* dan *False Negative* (fn) sebanyak 33 *record*. Berikutnya 67 *record* untuk *true negative* (tn) dan 42 *record* *False Negative* (fp). Sedangkan dari penelitian *accuracy* menggunakan algoritma *Naive Bayes* dapat kita lihat pada gambar 4 dengan grafik ROC dengan nilai AUC: 0.484 +/- 0.107 (mikro: 0.484) (*positive class: Positif*) sebagai berikut:



Sumber: Hasil Penelitian(2017)

Gambar 4. Kurva ROC Naive Bayes

Berdasarkan hasil percobaan penelitian yang dilakukan untuk memecahkan masalah sentiment etika posting dapat disimpulkan bahwa hasil uji coba menggunakan metode *naïve bayes* mempunyai tingkat akurasi sebesar 72,50% dan memiliki nilai AUC 0.484 +/- 0.107 (*mikro: 0.484*) (*positive class: Positif*).

Implikasi penelitian mengarahkan pada tiga aspek, yaitu: 1) Aspek Sistem merupakan Penerapan kebijakan dalam sentiment etika posting pada media sosial akan berpengaruh pada sistem, dengan melalui data-data yang ada akan mempermudah dalam menyaring sebuah berita. Dalam penerapannya dapat menggunakan *software* RapidMiner 5.3. yang merupakan *software open source*, sehingga bisa didapatkan dengan gratis. Dengan memanfaatkan perangkat yang ada. Seperti *hardware* dan jaringan, sistem akan berjalan dengan lebih baik dalam memecahkan masalah yang ada. 2) Aspek Manajerial merupakan manajemen kinerja seseorang dalam menyaring suatu postingan akan lebih membawa dampak yang signifikan. Selain itu penerapannya dapat dilakukan untuk penelitian berikutnya. 3) Penelitian Lanjutan merupakan penelitian semacam ini dapat dikembangkan pada bidang lain. Penelitian ini juga dapat dikembangkan dengan algoritma aturan klasifikasi yang lain seperti algoritma Support Vector Machine, Decision Tree, K-Nearest Neighbours, Neural Network dan lain-lain. Ataupun dengan menggunakan model akurasi yang lain seperti Genetic Algorithm atau Particle Swarm Optimization.

4. Kesimpulan

Pendekatan berbasis pembelajaran mesin terdapat perbedaan terhadap sentimen etika posting dikarenakan tidak diperlukan data *training* sehingga data yang ada mempunyai ketergantungan kepada kamus, jika kamus data yang ada lengkap maka performa sistem dalam menganalisa sentiment etika posting akan sangat baik hasilnya, berbeda dengan data yang tidak lengkap maka performa sistem dalam menganalisa sentiment etika posting akan buruk hasilnya. Untuk menghasilkan suatu algoritma dengan performa terbaik terhadap analisis sentimen posting dengan hasil akurasi yang baik maka diperlukan sebuah metode yang tepat.

Penelitian ini telah berhasil menerapkan model untuk melakukan klasifikasi sentiment etika posting di media sosial dengan model *Naïve Bayes*. Klasifikasi menggunakan *tools rapid miner* dengan *naïve bayes*. Dari penelitian yang dilakukan dalam pengujian model dengan menggunakan teknik mesin pembelajaran yaitu *Naive Bayes*(NB) itu akan diterapkan dengan menggunakan data *review* sentimen etika berinternet di media sosial dari 200 data *review* dalam teks bahasa Indonesia yang terdiri dari 100 *review* positif dan 100 *review* negatif. Model klasifikasi yang dihasilkan mendapatkan hasil *accuracy* pengujian klasifikasi dilakukan dengan menggunakan *Naive Bayes* (NB) yaitu mendapatkan nilai *accuracy* 72.50% dan AUC: 0.484 +/- 0.107 (*mikro: 0.484*) (*positive class: Positif*), maka dapat disimpulkan pengujian analisa sentiment dengan menggunakan *Naive Bayes* (NB) akan lebih baik hasilnya sehingga kita dapat memecahkan masalah analisa sentimen dengan memilih pemberitaan atau informasi di media sosial khususnya lebih akurat.

Referensi

- Darujati C, Gumelar AB. 2012. Pemanfaatan Teknik Supervised Untuk Klasifikasi Teks Bahasa Indonesia. J. LINK 16: 1–8.
- Kontopoulos E, Berberidis C, Dergiades T, Bassiliades N. 2013. Ontology-based sentiment analysis of twitter posts. Expert Syst. Appl. 40: 4065–4074.
- Ling J, Kencana I putu EN, Oka TB. 2014. Analisis Sentimen Menggunakan Metode Naïve Bayes Classifier Dengan Seleksi Fitur Chi Square. E-Jurnal Mat. 3: 92–99.
- Routray P, Kumar Swain C, Prava Mishra S. 2013. A Survey on Sentiment Analysis. Int. J. Comput. Appl. 76: 975–8887.
- Tripathy A, Agrawal A, Rath SK. 2015. Classification of Sentimental Reviews Using Machine

Learning Techniques. *Procedia Comput. Sci.* 57: 821–829.

Zhang Z, Ye Q, Zhang Z, Li Y. 2011. Sentiment classification of Internet restaurant reviews written in Cantonese. *Expert Syst. Appl.* 38: 7674–7682.